Euro. Jnl of Applied Mathematics (2018), *vol.* 29, *pp.* 685–707. © Cambridge University Press 2017
doi:10.1017/S095679251700033X

Self-exciting point process models for political conflict forecasting

N. JOHNSON, A. HITCHMAN, D. PHAN and L. SMITH

Department of Mathematics, California State University, Fullerton, CA, USA
emails: a.nathanjohnson@csu.fullerton.edu, annelise.hitchman@csu.fullerton.edu, dangphan89@csu.fullerton.edu, lausmith@fullerton.edu

(*Received 24 January 2017; revised 3 November 2017; accepted 3 November 2017; first published online 4 December 2017*)

In 2008, the Defense Advanced Research Project Agency commissioned a database known as the Integrated Crisis Early Warning System to serve as the foundation for models capable of detecting and predicting increases in political conflict worldwide. Such models, by signalling expected increases in political conflict, would help inform and prepare policymakers to react accordingly to conflict proliferation both domestically and internationally. Using data from the Integrated Crisis Early Warning System, we construct and test a self-exciting point process, or Hawkes process, model to describe and predict amounts of domestic, political conflict; we focus on Colombia and Venezuela as examples for this model. By comparing the accuracy of fitted models to the observed data, we find that we are able to closely describe occurrences of conflict in each country. Thus, using this model can allow policymakers to anticipate relative increases in the amount of domestic political conflict following major events.

Key words: Applications to social sciences, point processes, prediction theory

1 Introduction

On a global scale, humans practice societal dynamics that cause aggregation, organization, growth, and inevitable conflict; we see these trends unfold in countries worldwide [2]. When modelling or forecasting political conflict, issues arise not only due to the complexity behind human motivations/behaviours, but also the defining characteristics endemic to specific political regions [30]. For example, conflict dynamics for domestic issues, like the 2014 demonstrations in Ferguson, Missouri, involve different sets of actors, histories, and context than the dynamics of international conflict such as the 2011 Arab Spring [6]. Models to predict the occurrence or duration of these events rely on historical conflict data acquired from automated, open-source text processors [5].

In 2008, the Defense Advanced Research Projects Agency commissioned a programme known as the Integrated Crisis Early Warning System (ICEWS) [29]. Its purpose was to codify global political events into a database that could be used to create forecasting models for political conflict [17]. Models developed using this data set exhibit various degrees of accuracy/precision and can employ one or several models working in tandem to produce forecasting results [4, 10, 17].

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

Since the purpose of ICEWS was to create reliable forecasting models, it is important to find new methods and approaches to increase the geospatial and temporal precision of forecasting results. We propose an application of self-exciting point processes to describe and forecast conflict behaviour. We focus on two South American countries as examples for our application: Colombia and Venezuela. We chose these two countries specifically due to their relative proximity (to minimize the effect geography can produce on modelling results) and their differing governmental structures [20].

2 The data

We begin by introducing the data codification process used by ICEWS and then detail our extraction method to obtain two distinct datasets that contain the number of domestic, political conflict-events in Colombia and Venezuela from 1/1/2001–12/31/2013.

2.1 ICEWS data set

The events in the ICEWS data set are presented in two major aggregations. The first aggregation of data, denoted the Event Set, is created using an event coder, updated from the Text Annotation by Augmented Replacement Instructions, called JABARI [21]. Text from over 250 news sources globally is analysed based on natural language processing methods in order to provide characteristic data on each event [11]. These characteristic data include description of the political 'actor' in the event, the political 'target' in the event, the intensity of the event (denoted as a numeric value from negative 10 to positive 10 known as the Goldstein Score), and the location of the news source [11]. The type of event that took place is described using a Conflict and Mediation Event Observation code [10]. The creators of Conflict and Mediation Event Observation also aimed to make more specific descriptions regarding the type of event that took place [21]. Events with negative Goldstein scores denote conflict events, such as protest or war, while positive Goldstein scores imply peace-keeping events, such as surrender or promises for collaboration [11].

The second aggregation of data given by ICEWS is denoted the Ground Truth Set [30]. This set further analyses the ICEWS Event Set and contains fewer characteristics per event. These characteristics indicate event location as well as specific events of interest such as rebellion, insurgency, domestic political crisis, international political crisis, and ethnic/religious violence. Each of the categories is listed per month with a 1 (that type of event occurred that month) or a 0 (that type of event did not occur).

2.2 Selection of data

We focus on the South American region, namely Colombia and Venezuela, for the years 2001–2013 and select only events with negative intensities from the Event Set on a daily basis – this gives us a daily count of the number of conflict events in each country. To avoid introducing external forces/actors into our data set, we selected only internal conflict events for both countries. This was done by extracting events in which either Colombia was both the 'target country' and 'source country' or Venezuela was both the 'target country' and 'source country'. As such, our methods and results relate to the internal

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

conflict profiles of Colombia and Venezuela, respectively; some external influence arises in rare cases, however. For example, on March 3, 2008 (a Monday), Colombia experienced an unusually high number of internal conflict events. If we look at the historical context, we see that this is likely due to actions taken by both Ecuador and Venezuela to sever diplomatic ties with Colombia as the result of an incident over the prior weekend [28]. Days of this nature are rare in our data set, but the resulting high number of conflict events for these days is significant. Following these days, we may expect a period of heightened conflict due to the extreme event and we want to analyse any future trends based on major conflict. As a result, when modelling this phenomenon, we include all outlier data as meaningful information.

3 Previous works

Multiple forecasting models have been developed to predict political conflict based on statistical analysis and regression. We give brief synopses of these models before presenting previous applications of self-exciting point process models. Their successful implementation for crime and wartime casualty analysis serve as motivation for us to apply self-exciting point processes to model political conflict.

3.1 Political forecasting models

Mass political conflict has motivated various research stemming from political science and mathematical viewpoints in efforts to develop intra-national and international crisis forecasting models [1]. Early conflict models were implemented in the 1970s, including Singer and Wallace [24], Freeman and Job [9], and Andriole and Young [1]. These models loosely describe past events and contain few predictive measures, instead mostly serving as evaluative studies for in-sample fitting [30]. Under Defense Advanced Research Projects Agency, many organizations created models to forecast political conflict. Innovative Decisions, Inc. produced a Naive Bayesian network structure to combine a set of social science models [14]. Lockheed Martin also developed conflict forecasting models using ICEWS data; the company implemented six models including logistic regression models, Bayesian methods, geospatial network models, and agent-based models [3, 6, 19].

3.2 Self-exciting point process models

Self-exciting point processes were originally used to predict aftershock events from an earthquake. These models, denoted Epidemic-Type Aftershock Sequences models, are typically functions of time, space, and magnitude [15]. The models are created such that the likelihood of aftershock events decreases with distance and time from the earthquake epicenter [12]. Since then, various research used self-exciting point processes to model/forecast violence and terrorism. Lewis *et al.* applied a self-exciting point process to model the number of civilian deaths in Iraqi municipalities [13]. The model uses a Hawkes process which combines an assumed background rate, $\mu(t)$, with a self-exciting term that describes the effect past events can have on motivating future events [25]. They constructed $\mu(t)$ by performing a kernel density estimate to arrive at the background rate. With

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

this selection of $\mu(t)$, the model was able to describe relative increases in the point counts more accurately than the classical assumption of a constant background rate, μ_{const} [13].

Self-exciting point process models have also been applied to violence and crime in the United States. Egesdal *et al.* used a self-exciting point process to model gang rivalries in Los Angeles [8]. They simulated a Hawkes process, as well as a Poisson process, and found that the Hawkes process simulation produced a closer match to observed data. Mohler *et al.* used self-exciting point processes to model burglary data [15]. In this application, the self-exciting term was used to describe the effect that burglary success in one location had on the chance for a repeat burglary at that same or a nearby location. This model is currently used by law enforcement agencies for predictive policing [15].

In 2016, Tench *et al.* [26] found that self-exciting point process models can accurately fit data related to terrorist activity. They examined data of improvised explosive device (IED) attacks in Northern Ireland and extended a univariate self-exciting point process to a multi-dimensional model that accommodates realization of multiple adversaries in the self-exciting terms. Six models were constructed to describe interactions between various groups in Northern Ireland. For example, one such model described the interaction between events targeting British Security Forces and IED attacks in which Catholics were killed. Using the Kolmogorov–Smirnov test, Tench *et al.* find that some models provide adequate goodness-of-fit, while others are more susceptible to fringe or edge effects. Overall, however, they conclude that each self-exciting point process model closely follows trends in their data.

4 Methods and assumptions

We begin with a discussion regarding general point processes and self-exciting point processes. Section 4.2 outlines our model construction, while Section 4.3 discusses our methodology for parameter optimization.

4.1 General point processes

A point process is generally a random collection of points in some space [7]. Point processes typically count occurrences of a phenomenon (measured in non-negative integers, $\mathbb{Z}^+$) over a continuous or discrete space; each occurrence represents an event, and events may tend to occur with frequencies described by a random variable [18, 25]. For example, point processes can describe the number of customers entering an establishment as a function of time, or they can describe the number of earthquakes that occur in a particular region.

A common point process is the Poisson point process which assumes a constant rate of point occurrence, λ, such that

$$\Pr(N(t, t+\delta) = 1 | h_t) = \lambda\delta + \mathcal{O}(\delta^2),$$
$$\Pr(N(t, t+\delta) > 1 | h_t) = \mathcal{O}(\delta^2).$$

Here, $N(t, t+\delta)$ is the number of points in the time interval $(t, t+\delta)$, h_t is the history of the process up to time t, and $\delta > 0$ is arbitrarily small [7]. A core assumption of a general

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

Poisson process is that event occurrences within separate time intervals are considered to follow *independent* Poisson distributions. Thus, the probability of two events happening within an arbitrarily small time frame, $\delta > 0$, is assumed to be essentially zero. This allows for observed events to occur at distinct times, but the independence assumption removes the possibility that past events can influence future point occurrences.

4.2 Self-exciting point processes

Unlike the Poisson process, which assumes independence between time intervals, a self-exciting point process allows the distribution of past points to influence the future occurrence of events [25]. That is, we call a point process, N, self-exciting if

$$\mathrm{Cov}[N(t_1, t_2), N(t_2, t_3)] > 0$$

for all $0 < t_1 < t_2 < t_3$ [7].

If we have a point process which satisfies this condition, along with a given history, $h_t = \{t_1, t_2, \ldots, t_n\}$, then we can define the *conditional intensity function of N* as

$$\lambda(t|h_t) = \lim_{\delta \to 0} \frac{E[N(t, t+\delta)]}{\delta}.$$

Given the self-exciting assumption, we can implement a class of point processes, known as a linear Hawkes process, with the following general form in continuous time, $\lambda(t)$:

$$\lambda(t) = \mu(t) + \int_{-\infty}^{t} g(t-v) dN_v. \tag{4.1}$$

Here, $\mu(t)$ is the background rate of point occurrence (and hence, is non-negative), while $\int_{-\infty}^{t} g(t-v) dN_v$ is the self-exciting term. A common choice for the function $g(t - t_i)$ is

$$g(t - t_i) = \alpha e^{-\beta(t-t_i)}. \tag{4.2}$$

Yet, if a specific realization of N, h_t, is known, then an equivalent expression to equation (4.1) is

$$\lambda(t) = \mu(t) + \sum_{t_i < t} g(t - t_i). \tag{4.3}$$

In this case, the self-exciting term is now $\sum_{t_i<t} g(t - t_i)$. If we assume that a general, linear Hawkes process, N, is stable (i.e., does not grow without bound), then $\int_{\infty}^{t} g(t-v) dN_v$ must converge. For g of the form in equation (4.2), this occurs if $\frac{\alpha}{\beta} < 1$.

When considering the background rate, we have discretion over selecting what form it will take. Ozaki [16] and Snyder *et al.* [25] assume that the background rate is a constant, μ_{const}, while Lewis *et al.* [13] and Mohler *et al.* [15] conclude that a kernel-smoothed $\mu(t)$ provides a more reliable fit for their data sets. We consider two self-exciting point process models during our analysis, one with a constant background rate and one with a non-constant background rate. The implied interpretations of computed parameter values provides some insight as to which background rate provides a more accurate description of our data.

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

4.2.1 *The background rate*

We let N be a point process which counts the number of daily conflict events within a given region with expected value, $\lambda(t)$, at time t. We also let $\vec{y}$ be the data vector of daily conflict-event counts from 1/1/2001 to 12/31/2013 extracted from the ICEWS data set, making $\vec{y}$ a realization of N. We assume that daily event counts result from the sum of parent/prior events and offspring (self-excited) events. The expected value of this sum at time t is given by $\lambda(t)$ ($\lambda(t) \geqslant 0\ \forall t$), which is a function of continuous time. We assume $\lambda(t)$ is of the form in equation (4.3). We concentrate on two selections of our function, each a self-exciting point process that differ in their background rate. Let $\lambda_c(t)$ be the process with constant background rate μ_{const} and let $\lambda_k(t)$ be the process with non-constant background rate $\mu(t)$. Then, μ_{const} can be treated as a parameter of λ_c to be estimated at the optimization stage. The choice for the non-constant $\mu(t)$ is a kernel-smoothed function, dependent on the time of events.

The set $\{y_i\}$, $i \in \{1, 2, \ldots, 4749\}$, are the elements of our data vector, $\vec{y}$. Each y_i are the daily conflict-event counts on day i, and we assume that each conflict event in our data set occurs at a distinct time. This assumption is made to avoid uniformly spaced occurrence times. Thus, to generate a random distribution of arrival times, our conflict-event arrival time data takes the following form: for each day, i, we randomly select y_i times, $t_{i,j}$, $j \in \{1, 2, \ldots, y_i\}$, between $(i, i+1)$ as the conflict-event arrival times for that day. Then, for days in which multiple conflict events occur (i.e., $y_i > 1$), we have a set of arrival times, $t_i = \{t_{i,1}, t_{i,2}, \ldots, t_{i,y_i}\}$. As a result, our final conflict-event arrival time data set is

$$\tau = \{\{t_1\}, \{t_2\}, \ldots, \{t_n\}\}.$$

The set, τ, is sorted in ascending order and $|\tau| = \sum_{i=1}^{n} y_i$, so we have a distinct arrival time for each conflict event. Additionally, we know that at each arrival time in τ, the conflict-event count is equal to 1, so we can directly use equation (4.5). We take $n = 4,749$ (the total number of days from 1/1/2001 to 12/31/2013) points as training data for parameter optimization.

We construct our non-constant background rate, $\mu(t)$, using a Gaussian kernel, as

$$\mu(t) = \sum_{i=1}^{n} \sum_{j=1}^{y_i} K\left(\frac{t - t_{i,j}}{h}\right), \tag{4.4}$$

where

$$K\left(\frac{t - t_i}{h}\right) = \frac{1}{\sqrt{2\pi}h} e^{\frac{-(t-t_i)^2}{2h^2}}.$$

Additionally, we compute the bandwidth of $K(\frac{t-t_i}{h})$ as $h = (\frac{4\sigma_Y^5}{3n})^{1/5}$ using Silverman's Rule of Thumb with σ_Y as the sample standard deviation of the set of arrival times, τ [22,23]. We use Silverman's Rule of Thumb as the kernel bandwidth since it minimizes the mean integrated squared error given by

$$E\left[\int (\hat{f}_h(x) - f(x))^2 dx\right],$$

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

where

$$\hat{f}_h(x) = \frac{1}{hN}\sum_{i=1}^{N} K_h\left(\frac{x - x_i}{h}\right)$$

is a kernel-smoothed function based on a Gaussian kernel with bandwidth, h, and $f(x)$ is the estimated function from which known data are sampled. Silverman's Rule of Thumb is guaranteed to minimize the mean integrated squared error when the sampled data are normally distributed. Even though our conflict data are not normally distributed, Silverman's Rule of Thumb provides us with a variable bandwidth that can reduce the mean integrated squared error and allows us to avoid treating the kernel bandwidth as an additional unknown parameter [27].

4.2.2 *The self-exciting term*

We do not expect that the background rate influence will continue to affect future point magnitudes indefinitely. Offspring events may also generate their own offspring, so we restrict our selection of g to light-tailed distributions such that

$$\int_0^\infty g(t)dt < 1.$$

Otherwise, each prior event, would generate more than one offspring event, leading to a point process that would grow without bound. We proceed with g as an exponential distribution as in [8, 13, 15, 16]. Our self-exciting term is then

$$g(t - t_{i,j}) = \sum_{t_{i,j}<t} we^{-w(t-t_{i,j})}, \tag{4.5}$$

where w is the exponential distribution parameter and the $t_{i,j}$ are the elements of τ less than t. This is analogous to the self-exciting term in equation (4.2).

4.2.3 *The final model*

Equipped with two forms of the background rate as well as the self-exciting term, we have the following two rates of point occurrence:

$$\lambda_c(t) = \mu_{\text{const}} + \sum_{t_{i,j}<t} we^{-w(t-t_{i,j})},$$

$$\lambda_k(t) = \mu(t) + \sum_{t_{i,j}<t} we^{-w(t-t_{i,j})}.$$

Here, $\lambda_c(t)$ uses the constant background rate, μ_{const}, and $\lambda_k(t)$ uses the kernel -smoothed background rate, $\mu(t)$. To include information about the proportion of parent events and offspring events in $\lambda_k(t)$, we introduce a proportionality constant, p, and our final models become

$$\lambda_c(t) = \mu_{\text{const}} + k_0 \sum_{t_{i,j}<t} we^{-w(t-t_{i,j})}, \tag{4.6}$$

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

$$\lambda_k(t) = p\mu(t) + (1-p)\sum_{t_{i,j}<t} we^{-w(t-t_{i,j})}. \tag{4.7}$$

As such, p can be interpreted as the proportion of background events in $\lambda_k(t)$, and $1/w$ can be interpreted as the average amount of time between a self-excited event and the corresponding prior event in both $\lambda_c(t)$ and $\lambda_k(t)$. For $\lambda_c(t)$, k_0 represents the average number of offspring events resulting from prior events (background or self-excited events), while $(1-p)$ represents this for $\lambda_k(t)$ [13].

4.3 Optimization

We let $\vec{m}_c = (\mu_{\text{const}}, k_0, w)$ and $\vec{m}_k = (p, w)$ be the parameter vectors of $\lambda_c(t)$ and $\lambda_k(t)$, respectively. We can compute $\vec{m}_c$ and $\vec{m}_k$ using maximum likelihood estimation for each year of our data set (2001–2013). In Section 5, we present the resulting optimized parameters and interpret their values based on political dynamics in each country.

4.3.1 *Maximum likelihood estimation*

The log-likelihood function for a general point process, $\lambda(t)$, with parameter vector, $\vec{m}$, is

$$\ell(t, \vec{m}) = \sum_i \log(\lambda(t_i)) - \int_0^T \lambda(t)dt,$$

where T is the total timeframe of our data set [16]. For a Hawkes process of the form in equation (4.6) (constant background rate), this becomes

$$\ell(t, \vec{m}_c) = -\mu_{\text{const}}T + k_0\sum_{i=1}^{n}\sum_{j=1}^{y_i}\left[\left(e^{-w(T-t_{i,j})}-1\right)\right] + \sum_{i=1}^{n}\sum_{j=1}^{y_i}\log(\mu_{\text{const}} + k_0 A(i,j)), \tag{4.8}$$

where $A(i,j) = \sum_{t<t_{i,j}} we^{-w(t_{i,j}-t)}$ is summed over all arrival times less than $t_{i,j}$ [16]. For a Hawkes process of the form in equation (4.7) (non-constant background rate), the log-likelihood function is [25]

$$\begin{aligned}\ell(t, \vec{m}_k) = \int_0^T p\mu(s)ds + \sum_{i=1}^{n}\sum_{j=1}^{y_i}\left[(1-p)\left(e^{-w(T-t_{i,j})}-1\right)\right] \\ + \sum_{i=1}^{n}\sum_{j=1}^{y_i}\log(p\mu(t) + (1-p)A(i,j)).\end{aligned} \tag{4.9}$$

To maximize this function, a variety of numerical optimization routines can be used; we proceed with the Nelder–Mead simplex algorithm. We impose constraints on this optimization routine to ensure stability of the resulting model. From equations 4.2 and 4.7, we must have $k_0 < 1$ for λ_c and $(1-p) < 1$ for λ_k. Additionally, we must have $p \in (0,1)$ for λ_k. As an additional constraint on w, we assume that $\frac{1}{w} \in (0, 1/60)$.

We generate 500 random initial guesses for the Nelder–Mead algorithm within our constraints. Using these 500 initial guesses, we obtain 500 sets of parameters from the

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

Table 1. *Optimized parameters using maximum likelihood estimation for λ_c. Here, $\vec{m}_c = [\mu_{const}, k_0, w]$*

Year	Colombia $\vec{m}_c^{MLE} = [\mu_{\text{const}}\ (\text{day}^{-1}), k_0, w\ (\text{day}^{-1})]$	Venezuela $\vec{m}_c^{\text{MLE}} = [\mu_{\text{const}}\ (\text{day}^{-1}), k_0, w\ (\text{day}^{-1})]$
2001	$\vec{m}_c = [0.0113, 0.9997, 1.3853]$	$\vec{m}_c = [0.1550, 0.8958, 1.1915]$
2002	$\vec{m}_c = [0.0055, 1.0000, 1.2757]$	$\vec{m}_c = [0.0088, 1.0000, 1.4215]$
2003	$\vec{m}_c = [2.2169, 0.7603, 2.6931]$	$\vec{m}_c = [1.3284, 0.7250, 2.3947]$
2004	$\vec{m}_c = [0.1362, 0.9879, 1.7283]$	$\vec{m}_c = [0.4716, 0.9212, 1.6460]$
2005	$\vec{m}_c = [0.0800, 0.9916, 1.6189]$	$\vec{m}_c = [0.2258, 0.9679, 1.8250]$
2006	$\vec{m}_c = [0.0158, 0.9985, 1.3875]$	$\vec{m}_c = [0.0191, 0.9978, 1.5032]$
2007	$\vec{m}_c = [2.6426, 0.7247, 2.2555]$	$\vec{m}_c = [1.3781, 0.8270, 2.2853]$
2008	$\vec{m}_c = [1.9107, 0.8723, 2.4901]$	$\vec{m}_c = [1.4122, 0.7973, 2.5366]$
2009	$\vec{m}_c = [2.4249, 0.6727, 2.7340]$	$\vec{m}_c = [1.9843, 0.7127, 2.5374]$
2010	$\vec{m}_c = [1.4951, 0.7845, 2.5016]$	$\vec{m}_c = [1.2008, 0.8039, 2.2639]$
2011	$\vec{m}_c = [1.1864, 0.6721, 2.8686]$	$\vec{m}_c = [0.9618, 0.5550, 1.9932]$
2012	$\vec{m}_c = [1.1845, 0.6768, 2.1513]$	$\vec{m}_c = [1.1309, 0.6429, 1.7059]$
2013	$\vec{m}_c = [1.2849, 0.6907, 2.3021]$	$\vec{m}_c = [1.1813, 0.7928, 2.0032]$

optimization algorithm, and select the set which corresponds to the largest log-likelihood values. This is done for each model, λ_c and λ_k with respect to equations (4.8) and (4.9).

5 Results and discussion

Values of optimized parameters, m_c and m_k, are provided for each calendar year from 2001 through 2013. We then provide figures demonstrating the fit of $\lambda_k(t)$ to observed data for Colombia and Venezuela in select calendar years. Weekly average plots for both countries for each year are provided in the Appendix. Finally, we provide simulations compared to ground truth simulated data.

5.1 Optimized parameters for Colombia and Venezuela

Tables 1 and 2 compile the optimized parameters obtained using maximum likelihood estimation for λ_c and λ_k, respectively.

In Table 1, we see that the optimized parameters vary from year to year for each country, with μ_{const} generally increasing for later years in both countries. The values for μ_{const} range from 0.0055 (Colombia 2002) to 2.6426 (Colombia 2007). In many of the years with smaller μ_{const} values, there are gaps consisting of many days between events (see Appendix figures), resulting in a series of 0 event counts. Thus, given these data, we expect the background rates to be lower in these years. We note that for all years without major gaps, μ_{const} for Venezuela is lower than for Colombia; this may indicate that self-excitation was relatively more prevalent in the Venezuela data. For all years, $w > 1$, indicating the average time between a self-excited event and its corresponding prior event is less than one day. For both countries, when $\mu_{\text{const}} > 1.18$, this time is less than half a day.

Allowing a constant background rate, like μ_{const}, to represent an entire year causes the model to lose accuracy/relevance, especially when there are large gaps of time between

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

Table 2. *Optimized parameters using maximum likelihood estimation for λ_k. Here, $\vec{m}_k = [p, w]$*

Year	Colombia $\vec{m}_k^{\text{MLE}} = [p, w\ (\text{day}^{-1})]$	Venezuela $\vec{m}_k^{\text{MLE}} = [p, w\ (\text{day}^{-1})]$
2001	$\vec{m}_k = [0.33, 2.72]$	$\vec{m}_k = [0.48, 3.40]$
2002	$\vec{m}_k = [0.39, 2.78]$	$\vec{m}_k = [0.29, 2.58]$
2003	$\vec{m}_k = [0.34, 3.03]$	$\vec{m}_k = [0.36, 2.38]$
2004	$\vec{m}_k = [0.28, 2.74]$	$\vec{m}_k = [0.23, 2.02]$
2005	$\vec{m}_k = [0.31, 3.15]$	$\vec{m}_k = [0.29, 3.08]$
2006	$\vec{m}_k = [0.32, 2.52]$	$\vec{m}_k = [0.27, 2.47]$
2007	$\vec{m}_k = [0.33, 2.53]$	$\vec{m}_k = [0.27, 2.37]$
2008	$\vec{m}_k = [0.17, 2.55]$	$\vec{m}_k = [0.25, 2.92]$
2009	$\vec{m}_k = [0.39, 2.90]$	$\vec{m}_k = [0.34, 2.56]$
2010	$\vec{m}_k = [0.36, 2.83]$	$\vec{m}_k = [0.32, 2.52]$
2011	$\vec{m}_k = [0.39, 3.04]$	$\vec{m}_k = [0.53, 2.23]$
2012	$\vec{m}_k = [0.44, 2.51]$	$\vec{m}_k = [0.55, 3.01]$
2013	$\vec{m}_k = [0.40, 2.68]$	$\vec{m}_k = [0.33, 2.38]$

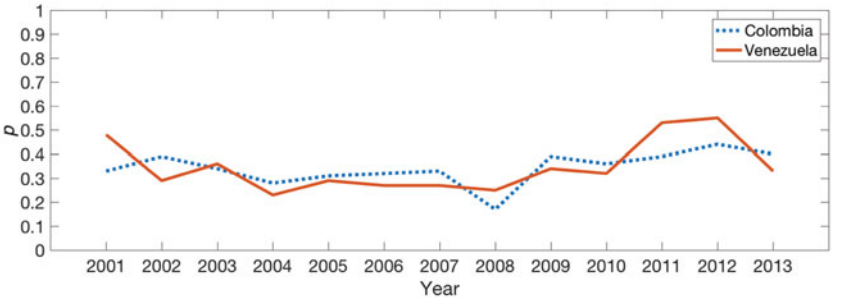


FIGURE 1. Maximum likelihood estimate for p for each year in Colombia and Venezuela.

events. Therefore, it seems more appropriate to use λ_k over λ_c due to the way λ_c treats the background rate – under λ_c, any increase in conflict beyond μ_{const} is attributed solely to self-excitation. This neglects the complexities of conflict involving political climate, historical context, and human motivation. Reducing independent conflict to a uniform measure in this manner does not accurately reflect historical experience with large-scale conflict and as such can be discounted. Alternatively, representing the background rate as a function of time allows the model to accommodate fluctuating conflict trends within a geospatial region.

Similar to the model for λ_c, the optimized parameters for λ_k vary from year to year and vary from country to country. Figure 1 shows how the optimal value for p changes for each year. To help understand why there are differences among years in the proportion of background events, p, the data for both countries are displayed for the years 2008 and 2011 in Section 5.2. In 2008, there was a day with over 300 events (March 3) in Colombia, so we might expect the proportion of self-excitation to be higher in this year than for other years. On this day, both Ecuador and Venezuela severed diplomatic

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

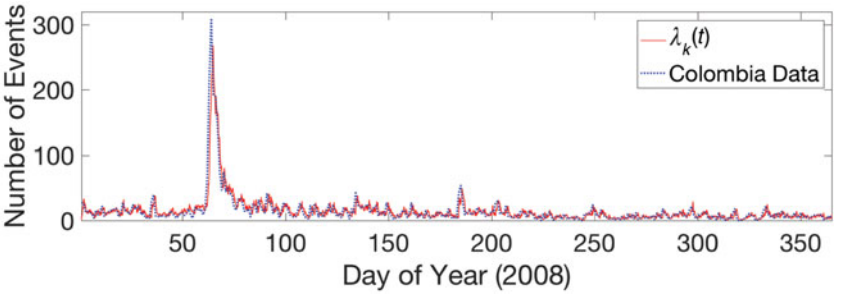


FIGURE 2. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2008. Here, $m_k = [0.17, 2.55]$.

ties with Colombia in an event known as the Andean Diplomatic Crisis. During the previous weekend, Colombian military units entered Ecuador to fight members of the Revolutionary Armed Forces of Colombia, a guerrilla group embroiled in ongoing civil conflict with Columbia. This caused Ecuador and Venezuela to dismantle their embassies in Bogota.

Accordingly, we see a decline in the value of p in 2008 for both the Colombia and Venezuela data when compared with previous years. From Figures 2 and 4 below, sharp increases in the number of daily conflict events occur for both countries on the same day with Venezuela experiencing around 100 events in a single day.

The parameter, p, may serve as an indicator for trends in internal conflict among both Colombia and Venezuela. Based solely on the daily number of conflict events in a given region, the parameter p may summarize relationships between time and amounts of conflict. In this case, it seems to indicate that Colombia experienced approximately two years, 2004 and 2008, with relatively high amounts of self-excitation ($p \leqslant 0.3$) over our data set, while Venezuela experienced approximately six such years (2002, 2004, 2005, 2006, 2007, 2008).

We also note that, for λ_k, optimal values for w range between 2.02 day^{-1} (Venezuela 2004) and 3.4 day^{-1} (Venezuela 2001). This means that the average time between a self-excited event and its corresponding prior event is less than half a day – this is consistent with results obtained for λ_c.

5.2 Model fit

As an example of the closeness of $\lambda_k(t;\vec{m}_k)$ to our data, Figures 2 and 3 show plots of $\lambda_k(t;\vec{m}_k)$ against the data for Colombia for calendar years 2008 and 2011. The model closely follows the data for each year. Note that the parameter values for each year are quite different, but this difference allows for the model to capture the greatly varying behaviour. In 2008, there was a day with over 300 conflict events, but 2011 only had days with less than 40 conflict events. In the appendix, Figures A 8 and A 11 display the model and the data in bins of 7-day averages for better visualization and to demonstrate the model's ability to closely follow trends in the data.

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

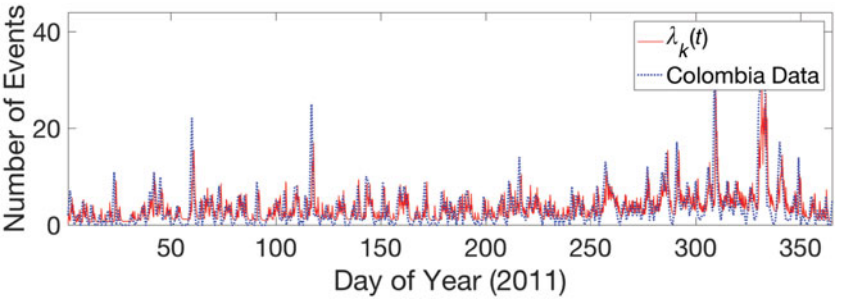


FIGURE 3. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2011. Here, $m_k = [0.39, 3.04]$.

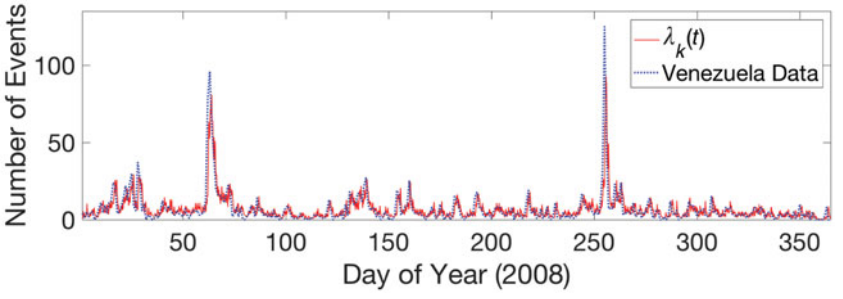


FIGURE 4. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2008. Here, $m_k = [0.25, 2.92]$.

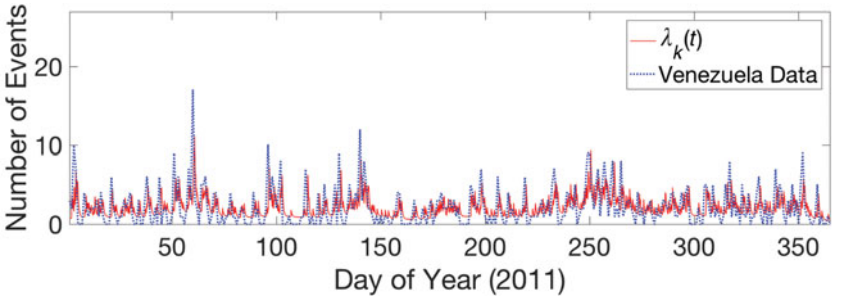


FIGURE 5. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2011. Here, $m_k = [0.53, 2.23]$.

To compare to Venezuela, Figures 4 and 5 show plots of $\lambda_k(t;\vec{m}_k)$ for calendar years 2008 and 2011. The model closely follows the data for each year.

To further compare the model fits, we look at the Akaike Information Criterion scores for λ_k, λ_c, and for $\lambda(t) = \mu(t)$. A lower score indicates a better model fit. The last of these, $\mu(t)$, is used to compare to a model with no self-excitation. The results for the Colombia yearly data are presented in Figure 6. To help see the differences in models, we have plotted the difference between each of these and λ_k, which has the lowest Akaike Information Criterion score for all years. This plot is given in Figure 7. In the year 2008,

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

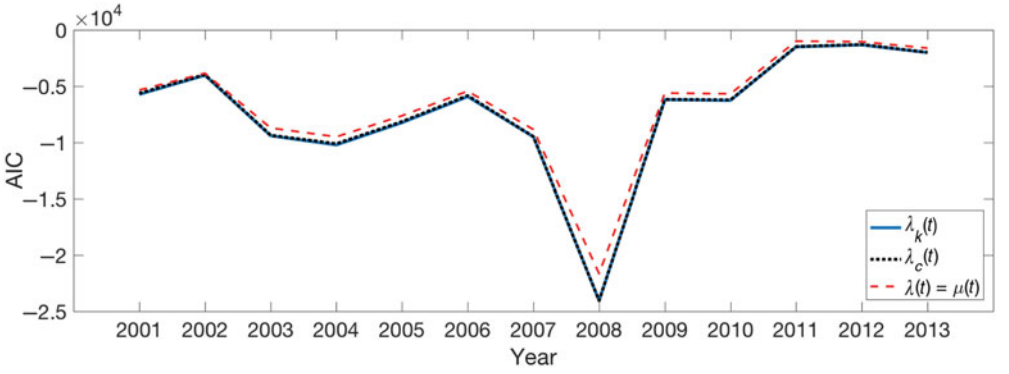


FIGURE 6. AIC scores for λ_k, λ_c, and $\mu(t)$ according to each calendar year in Colombia.

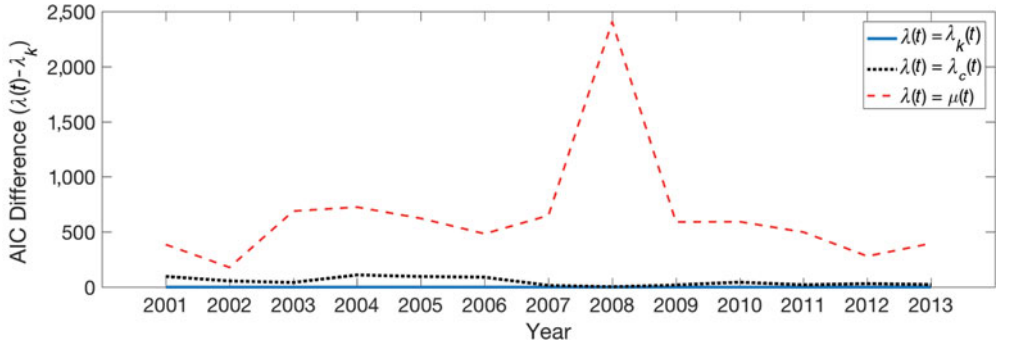


FIGURE 7. Difference in AIC scores from that of λ_k for λ_k, λ_c, and $\mu(t)$ according to each calendar year in Colombia.

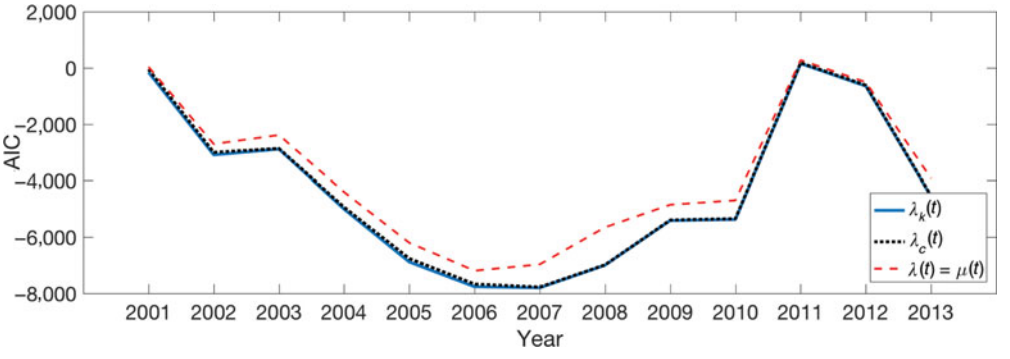


FIGURE 8. AIC scores for λ_k, λ_c, and $\mu(t)$ according to each calendar year in Venezuela.

where the value of p significantly dropped for the model λ_k, we see a significant increase in the Akaike Information Criterion difference for $\lambda(t) = \mu(t)$. The model fits are similar for Venezuela, shown in Figures 8 and 9. For each year, the model λ_k has the lowest Akaike Information Criterion score of all three models.

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

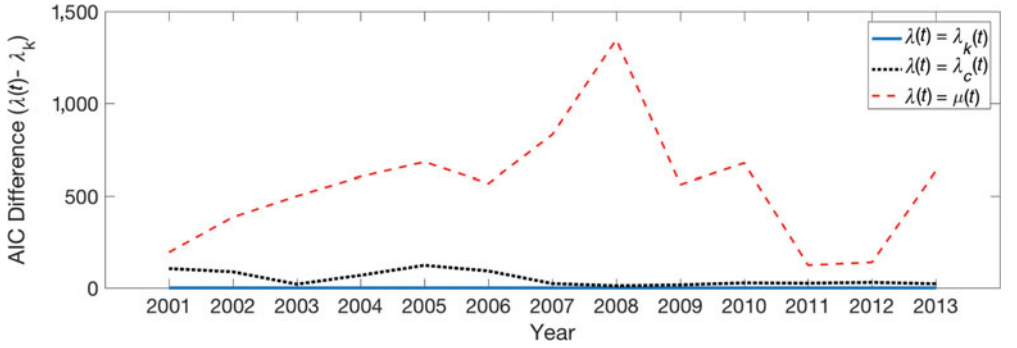


FIGURE 9. Difference in AIC scores from that of λ_k for λ_k, λ_c, and $\mu(t)$ according to each calendar year in Venezuela.

6 Conclusions and implications

Self-exciting point process models are able to accurately describe the amount of political conflict in a given region. We constructed two models, λ_c and λ_k, for conflict in both Colombia and Venezuela over the period 1/1/2001–12/31/2013. The major difference between these two models is the construction of the background rate; λ_c uses a constant background rate while λ_k uses a kernel-smoothed background rate based on a normal kernel. We found that a kernel-smoothed background rate not only provides a better fit to data, but also reflects realistic considerations of conflict such as political context.

Using maximum likelihood estimation, we optimized the parameters for each model on a yearly basis. We noticed that the parameter values changed from year to year, providing insight into a country's conflict dynamics. The proportion of background events in Colombia was generally lower than the proportion of background events in Venezuela. This is perhaps due to the fact that outliers occur more frequently in Venezuela than in Colombia. As a result, models for Colombian internal conflict are more sensitive to self-excitation. Models for Venezuelan conflict show that Venezuela generally has high levels of internal conflict, causing high estimations of background proportions. Additionally, we found that the average amount of time between a self-excited event and its corresponding prior event is less than a day for both countries. This is true regardless of using a constant background rate or a kernel-smoothed background rate. As a result, our self-exciting point process models imply that responses to major political conflict events will occur quickly.

By computing sets of optimal parameters for each year, we are able to trace differences through a country's history. We can make especially clear comparisons between the values of p (the proportion of background events to total events) not only between Colombia and Venezuela, but also between different years within a single country. As an example, our computed parameters suggest that levels of self-excitation increased in Venezuela following the death of President Hugo Chavez in 2013 when compared to prior years. The Venezuelan proportion of self-excitation in 2013 is approximately 22% higher than in 2012. Our models facilitate additional comparisons of this nature and may be helpful to inform policy making by providing additional context to political conflict.

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

Overall, the proposed self-exciting point process model, λ_k, was able to capture the nature of a country's internal conflict events given only the number of events per day.

7 Future work

There are several possible avenues to build on this work. We focused on looking at the internal conflict within a single country. An interesting extension would be to incorporate space and distance as a factor, while examining conflict between countries.

Further, our self-exciting term relied on the exponential distribution due to its small number of parameters. This assumption, however, means that the effect of self-excitation has high magnitude and occurs immediately after an independent conflict event. It is perhaps intuitive to assume that the effect of self-excitation maximizes sometime after a background event before beginning to decay. As a result, replacing the self-exciting term with a distribution following this property, such as a gamma distribution, may produce a more accurate fit to data.

Acknowledgements

We would like to thank Angela O'Mahoney, Osonde Osoba, and Bill Welser from the RAND Corporation for providing us the ICEWS data and acknowledge their feedback regarding the model's construction. This paper and all research on which it relies were not produced by and do not represent the work of the RAND Corporation.

References

[1] Andriole, S. & Young, R. (1977) Toward the development of an integrated crisis early warning system. In: *International Studies Quarterly*, pp. 107–50.

[2] Atran, S. & Ginges, J. (2012) Religious and sacred imperatives in human conflict. *Science* **336**(6083), 855–57.

[3] Barathy, G. & Silverman, B. (2012) Holistically evaluation agent-based social systems models: A case study. *Simulation* **89**(1), 102–135.

[4] Bond, D., Jenkins, J., Taylor, C. & Schock, K. (1997) Mapping mass political conflict and civil society issues and prospects for the automated development of event data. *J. Conflict Resolution* **41**(4), 553–579.

[5] Boshcee, E., Natarajan, P. & Weischedel, R. (2013) Automatic extraction of events from open source text for predictive forecasting. In: *Handbook of Computational Approaches to Counterterrorism*, pp. 51–67.

[6] Brandt, P., Freeman, J. & Schrodt, P. (2011) Real time, time series forecasting of inter-and-intra-state political conflict. In: *Conflict Management and Peace Science*, pp. 41–64.

[7] Cox, D. & Isham, V. (1980) *Point Processes*, CRC Press, London.

[8] Egesdal, M., Fathauer, C., Jouie, K., Neuman, J., Mohler, G. & Lewis, E. (2010) Statistical and stochastic modeling of gang rivalries in Los Angeles. *SIAM Undergraduate Research Online* 72–94.

[9] Freeman, J. & Job, B. (1979) Scientific forecasts in international relations: Problems of definition and epistemology. *International Studies Quarterly* **23**(1), 113–43.

[10] Germer, D., Schrodt, P., Yilmaz, O. & Abu-Jabr, R. (2002) Conflict and mediation event observation (CAMEO): A new event data framework for the analysis of foreign policy interactions. *International Studies Association.*

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

[11] Goldstein, J. (1992) A conflict-cooperation scale for weis events data. *J. Conflict Resolution* **36**(2), 369–385.
[12] Helmstetter, A. & Sornetter, D. (2003) Importance of direct and indirect triggered seismicity in the etas model of seismicity. *Geophys. Res. Lett.* **30**(11), 1576–1579.
[13] Lewis, E., Mohler, G., Brantingham, P. & Bertozzi, A. (2012) Self-exciting point process models of civilian deaths in Iraq. *Secur. J.* **25**(3), 244–264.
[14] Mahoney, S., Comstock, E. & Darcy, S. (2011) Aggregating forecasts using a learned bayesian network. In: *Twenty-Fourth International FLAIRS Conference.*
[15] Mohler, G., Short, M., Brantingham, P., Schoenberg, F. & Tita, G. (2012) Self-exciting point process modeling of crime. *J. Am. Stat. Assoc.* **106**(493), 100–108.
[16] Ozaki, T. (1979) Maximum likelihood estimation of Hawkes' self-exciting point processes. *Ann. Inst. Stat. Math.* **31**(1) 145–155.
[17] O'Brien, S. (2013) A multi-method approach for near real time conflict and crisis early warning. In: *Handbook of Computational Approaches to Counterterrorism* pp. 401–418.
[18] Perkel, D., Gerstein, G. & Moore, G. (1967) Neuronal spike trains and stochastic point processes: the single spike train. *Biophys. J.*, **7**(4), 391–418.
[19] Racette, M., Smith, C., Cunningham, M,. Heekin, T., Lemley, J. & Mathieu, R. (2014) Improving situational awareness for humanitarian logistics through predictive modeling. In: *Systems and Information Engineering Design Symposium (SIEDS)*, pp. 334–339.
[20] Roberts, K. (2003) Social correlates of party system demise and populist resurgence in Venezuela. *Latin America Politics and Society* **45**(3), 35–57.
[21] Schrodt, P., Yilmaz, O., Gerner, D. & Hermreck, D. (2008) The CAMEO (conflict and mediation event observations) actor coding framework. *2008 Annual Meeting of the International Studies Association.*
[22] Sheather, S.(2004) Density estimation. *Statistical Science* **19**(4), 588–597.
[23] Silverman, B.(1986) *Density Estimation for Statistics and Data Analysis*, CRC Press, London.
[24] Singer, J. & Wallace, M. (1979) To augur well: Early warning indicators in world politics. *Sage Publications*, First edition.
[25] Snyder, D. & Miller, M. (1991) Self-exciting point processes. Second edition. *Random Point Processes in Time and Space*, 287–340.
[26] Tench, S., Fry, H. & Gill, P. (2016) Spatio-temporal patterns of IED usage by the provisional Irish Republican Army. *Eur. J. Appl. Math.* **27**(3), 377–402.
[27] Terrell, G. & Scott, D. (1992) Variable kernel density estimation. *Ann. Stat.* 1236–1265.
[28] Waisberg, T.2008 Colombia's use of force in Ecuador against terrorist organization: International law and the use of force against non-state actors. *Am. Soc. Int. Law.* **12**(17).
[29] Ward, M., Beger, A., Cutler, J., Dickenson, M., Dorff, C. & Radford, B. (2013) Comparing GDELT and ICEWS event data. *Analysis*, **21**(1) 267–97.
[30] Ward, M., Metternich, N., Carrington, C., Dorff, C., Gallop, M., Hollenbach, F., Schultz, A. & Weschle, S. (2012) Geographical models of crises: Evidence from ICEWS. *Adv. Design Cross-Cultural Act.* 429.

Appendix A Plots for Colombia for 2001–2013

Figures A 1–A 13 give the model fits for the years 2001–2013 in Colombia. The solid red curves give the average weekly model value for $\lambda_k(t)$, using the kernel-smoothed background rate. The blue dashed curves give the weekly average number of conflict events recorded per day. Note that in all years, the model closely matches the data.

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

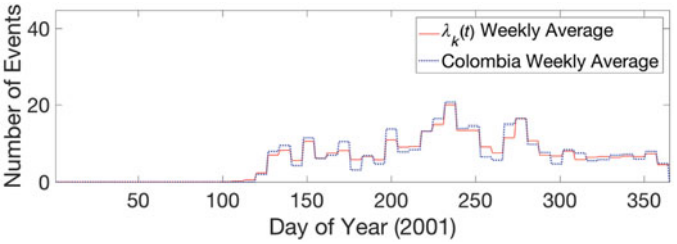


FIGURE A 1. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2001.

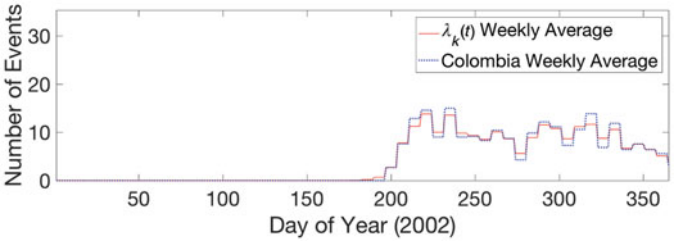


FIGURE A 2. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2002.

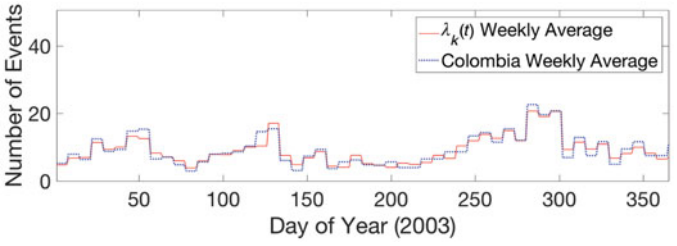


FIGURE A 3. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2003.

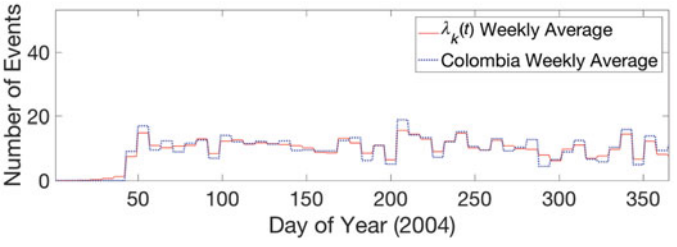


FIGURE A 4. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2004.

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

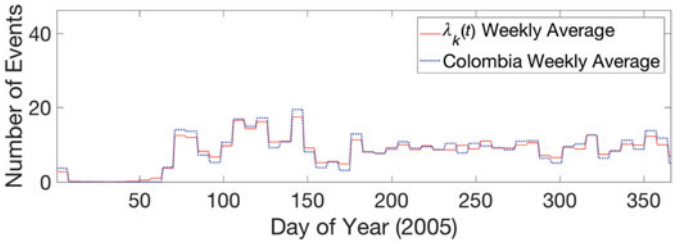


FIGURE A 5. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2005.

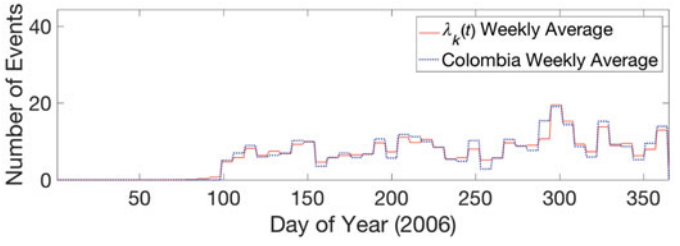


FIGURE A 6. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2006.

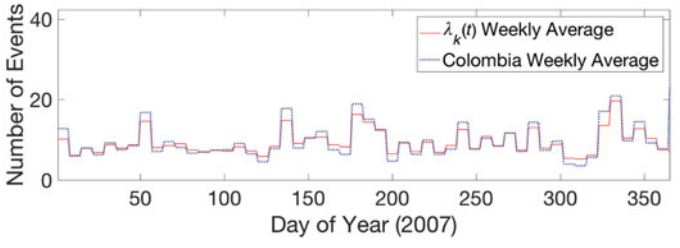


FIGURE A 7. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2007.

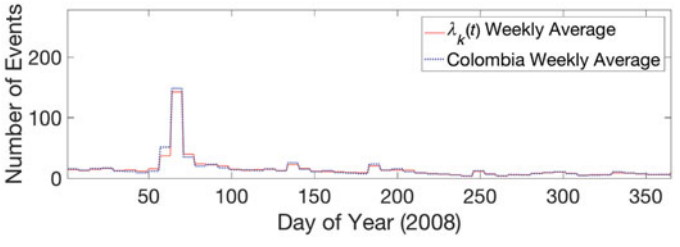


FIGURE A 8. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2008.

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

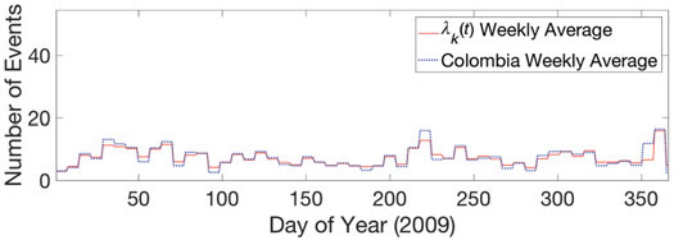


FIGURE A 9. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2009.

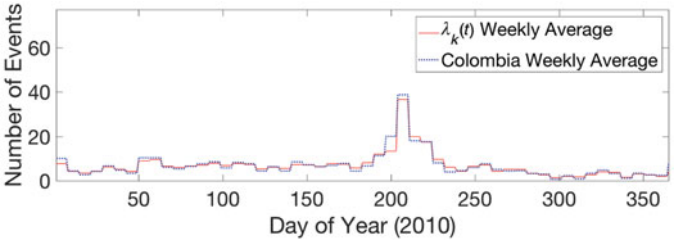


FIGURE A 10. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2010.

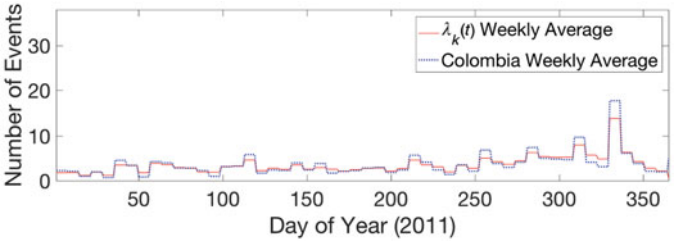


FIGURE A 11. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2011.

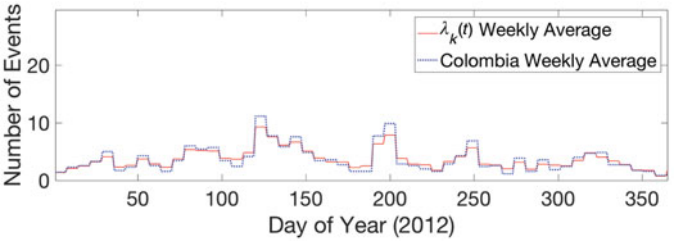


FIGURE A 12. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2012.

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

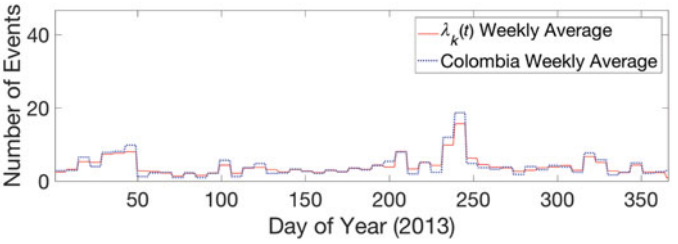


FIGURE A 13. Kernel-smoothed self-exciting point process model with Colombia data for the calendar year 2013.

Appendix B Plots for Venezuela for 2001–2013

Figures B 1–B 13 give the model fits for the years 2001–2013 in Venezuela. The solid red curves give the average weekly model value for $\lambda_k(t)$, using the kernel-smoothed background rate. The blue dashed curves give the weekly average number of conflict events recorded per day. Note that in all years, the model closely matches the data.

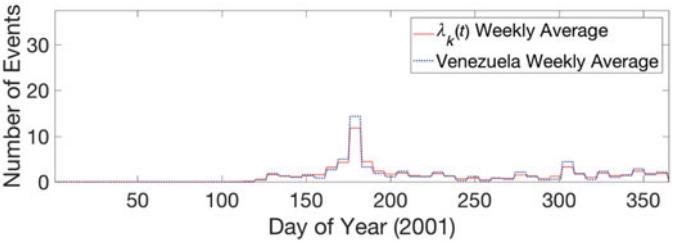


FIGURE B 1. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2001.

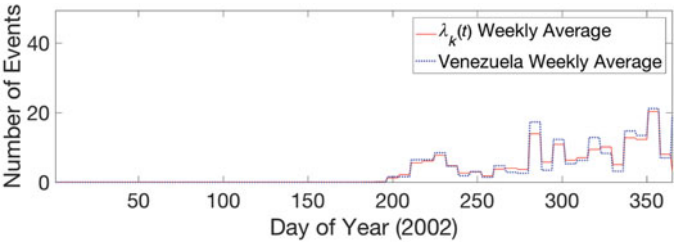


FIGURE B 2. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2002.

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

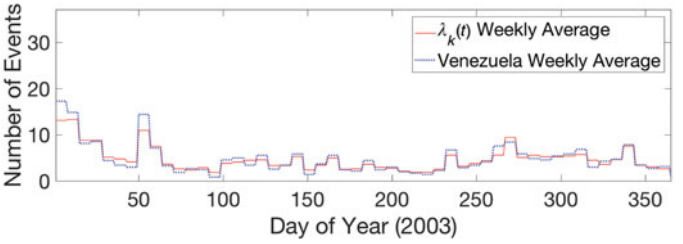


FIGURE B 3. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2003.

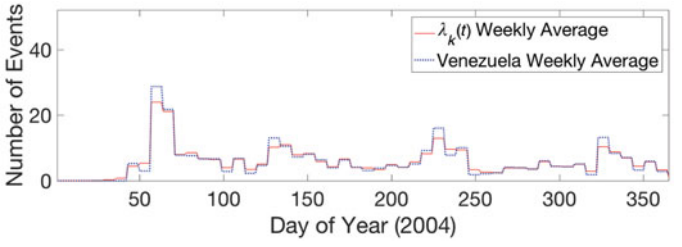


FIGURE B 4. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2004.

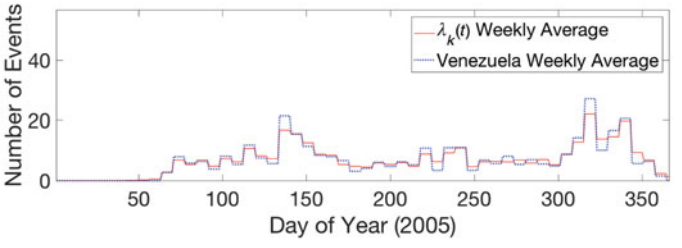


FIGURE B 5. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2005.

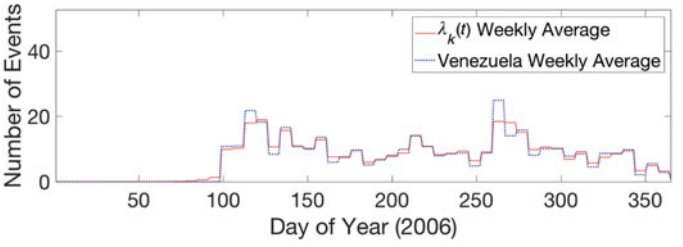


FIGURE B 6. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2006.

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

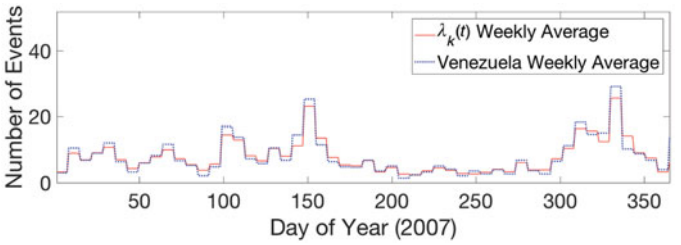


FIGURE B 7. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2007.

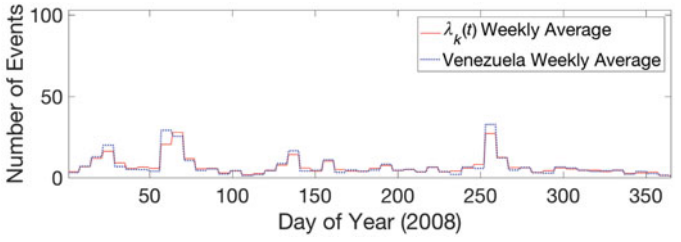


FIGURE B 8. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2008.

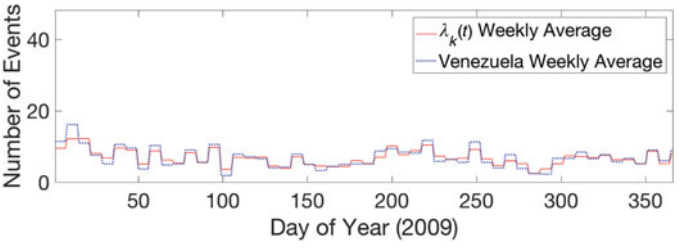


FIGURE B 9. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2009.

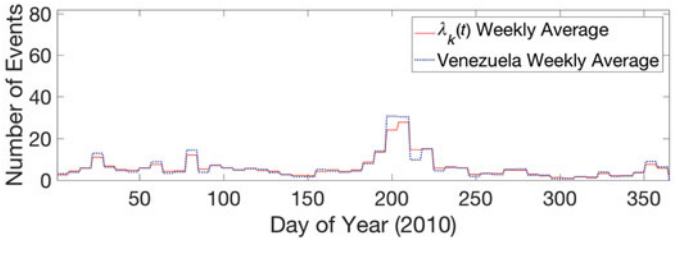


FIGURE B 10. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2010.

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press

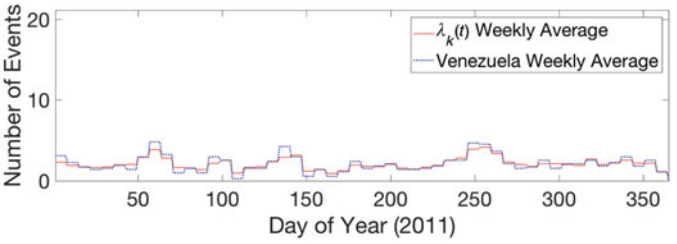


FIGURE B 11. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2011.

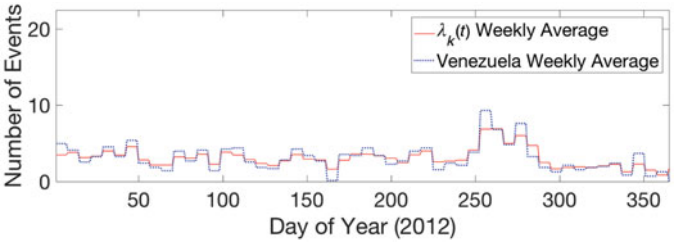


FIGURE B 12. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2012.

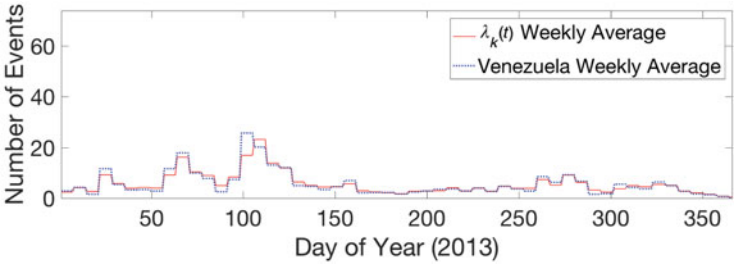


FIGURE B 13. Kernel-smoothed self-exciting point process model with Venezuela data for the calendar year 2013.

https://doi.org/10.1017/S095679251700033X Published online by Cambridge University Press